A High-Dimensional Nonparametric Multivariate Test for Mean Vector

Lan Wang, Bo Peng and Runze Li

November 25, 2013

Abstract

This work is concerned with testing the population mean vector of nonnormal high-dimensional multivariate data. Several tests for high-dimensional mean vector have been proposed in the literature, but they may not perform well for high-dimensional continuous, nonnormal multivariate data, which frequently arise in genomics studies and quantitative finance. This paper aims to develop a novel high-dimensional nonparametric test for the population mean vector so that multivariate normality assumption becomes unnecessary. With the aid of new tools in modern probability theory, we proved that the limiting null distribution of the proposed test is normal under mild conditions for $p > n$. We further study the local power of the proposed test and compare its relative efficiency with a modified Hotelling T^2 test for high-dimensional data. Our theoretical results indicate that the newly proposed test can have even more substantial power gain than the traditional nonparametric multivariate test does with finite fixed p. We assess the finite sample performance of the proposed test by examining its size and power via Monte Carlo studies. We illustrate the application of the proposed test by an empirical analysis of a genomics data set.

KEY WORDS: Asymptotic relative efficiency; High dimensional multivariate data; Hotelling T^2 test; Nonparametric test.

[1]Lan Wang is Associate Professor and Bo Peng is graduate student, School of Statistics, University of Minnesota, Minneapolis, MN 55455. Email: wangx346@umn.edu. Runze Li is Distinguished Professor, Department of Statistics and the Methodology Center, the Pennsylvania State University, University Park, PA 16802-2111. Email: rzli@psu.edu. Wang and Peng's research is supported by a NSF grant DMS1308960. Li's research is supported by NIDA, NIH grants P50 DA10075 and P50 DA036107. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

1 Introduction

Testing a hypothesis on the population mean is of fundamental importance in the statistical literature. It becomes very challenging for high-dimensional multivariate data since the traditional Hotelling T^2 test for mean vector is not well defined as the inverse of sample covariance matrix may not exist for the large p, small n problem. Here and throughout this paper, p stands for the number of variables (or features) of the data, and n for the sample size. It has been observed in Bai and Saranadasa (1996) that the power of the Hotelling T^2 test can be adversely affected even when $p < n$, if the sample covariance matrix is nearly singular. Several extensions of Hotelling T^2 test to high-dimension with a general covariance matrix have been proposed in the literature, see Bai and Saranadasa (1996) for the setting with $p/n \to c \in (0, 1)$, Srivastava and Du (2008) for the setting with $n = O(p^\delta)$ for some $1/2 < \delta \leq 1$ and Chen and Qin (2010) for the setting with the assumption $\mathrm{Tr}(\Sigma^4) = o((\mathrm{Tr}(\Sigma^2))^2)$, an implicit condition on the relationship between p and n, and among others.

This work was motivated from an empirical analysis of a microarray data set, for which the marginal distributions of microarray expressions seem to be nonnormal and heavy tailed based on values of their marginal kurtosises (Section 3.2). This leads us to develop a nonparametric test for high-dimensional population mean vector or the location parameter without multivariate normality assumption. Specifically, let $X_1, \ldots, X_n$ be an independent and identically distributed (iid) p-dimensional random sample from the model $X_i = \mu + \epsilon_i$, where ϵ_i is the random error to be specified later. In this paper, we consider a novel nonparametric procedure for testing the hypothesis

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0, \tag{1}$$

when p is potentially much larger than n.

We propose a new nonparametric test for hypothesis (1) based on spatial signs of the observations, and further study the asymptotic theory of the new test. Comparing with the extensions of Hotelling's T^2 test (Chen and Qin, 2010), the theory for the nonparametric test with $p > n$ is considerably more challenging. To derive the asymptotic theory, we employ new probability tools on the concentration properties of certain quadratic forms, which may be of independent interest and have potential applications in developing the theory for other related high-dimensional nonparametric procedures. The proposed nonparametric test has several appealing properties. First it is directly applicable for the setting with $p > n$, and it is computationally simple. Second, the new test is shown to lose little efficiency when the underlying data are multivariate normal and to have potential significant efficiency gain for heavy-tailed multivariate distributions. This is verified by deriving its asymptotic relative efficiency. From our Monte Carlo simulation, significant efficiency gain can be achieved at moderate sample size.

Nonparametric statistical procedures have been explored little in the high dimensional setting. This work takes a substantial step towards understanding their properties when $p > n$. Our theoretical analysis reveals a striking phenomenon: the efficiency gain of the new nonparametric test in the high-dimensional setting can be more substantial comparing with the well known traditional nonparametric tests efficiency gain in the "classical" framework where p is fixed and n goes to infinity. For example, consider the p-dimensional multivariate t-distribution with 3 degrees of freedom, which is heavy-tailed. For this distribution, it is well known that the asymptotic relative efficiency of the spatial sign test versus Hotelling's T^2 test is 1.9 for $p = 1$, 2.02 for $p = 3$, and 2.09 for $p = 10$. This implies an increasing trend as the dimension p

increases. The theory established in this paper suggests that when $p > n$, the asymptotic relative efficiency of the proposed new nonparametric test versus Chen and Qin's extension of Hotelling's T^2 test is about 2.54. This result provides strong support for the usefulness of nonparametric tests in high-dimensional problems.

The rest of the paper is organized as follows. In Section 2, we introduce the high-dimensional nonparametric test. We further derive its limiting null distribution under a set of weak conditions in Section 2.1, and investigate its power performance under local alternatives and study the asymptotic relative efficiency in Section 2.2. We conduct Monte Carlo simulations and further demonstrate the new test by analyzing the gene sets from a genomics study in Section 3. Some useful lemmas and technical proofs are given in the Appendix.

2 A high-dimensional nonparametric test

Throughout this paper, it is assumed that the random vector X_i follows a p-dimensional elliptical distribution. Elliptical distribution has been well studied in the statistical literature (Fang, Kotz and Ng, 1990), and has been considered to be useful models for finance data (McNeil, Frey and Embrechts, 2005). For an elliptically distributed random variable X_i, one has the following stochastic representation:

$$X_i = \mu + \epsilon_i, \quad \text{and} \quad \epsilon_i = \Gamma R_i U_i, \tag{2}$$

where Γ is a $p \times p$ matrix, U_i is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, and R_i is a nonnegative random variable independent of U_i. It has been shown that the distribution of X_i depends on Γ only through $\Gamma\Gamma^T$ (Fang, Kotz and Ng, 1989). Thus, denote $\Omega = \Gamma\Gamma^T$ for easy future reference. Note that if X_i is a multivariate normal vector with mean μ and covariance matrix Σ, then X_i can be expressed in the form of (2) with R_i^2 having a chi-square distribution with p degrees of freedom and $\Omega = \Sigma$. In general, under (2) the population covariance matrix Σ is related to Ω by $\Sigma = p^{-1}E(R_i^2)\Omega$.

As shown in Chapter 3 of Fang, Kotz and Ng (1990), the class of elliptical distributions is a rich family which includes multivariate normal distribution, multivariate t distribution, multivariate logistic distribution, Kotz-type multivariate distribution, and Pearson II type multivariate distribution and many others. Elliptical distribution provides a very flexible extension of multivariate normality and encompasses many useful heavy-tailed multivariate distributions. Furthermore, the class of elliptical distributions is recognized for its ability to accommodate tail dependence (the phenomenon of simultaneous extremes), which is not allowed by the multivariate normal distribution (Schmidt, 2002), in the research area of quantitative finance.

Our test statistic T_n is based on the spatial sign function of the observed data. Denote by $\|X_i\|$ the L_2 norm of X_i, the spatial sign function of X_i is defined as

$$Z_i = \begin{cases} \frac{X_i}{\|X_i\|}, & X_i \neq 0, \\ 0, & X_i = 0. \end{cases} \tag{3}$$

The spatial sign vector is simply the unit vector in the direction of X_i. In the univariate case, it reduces to the familiar sign function.

We propose the following new nonparametric test for (1):

$$T_n = \sum_{i=1}^{n} \sum_{j=1}^{i-1} Z_i^T Z_j. \tag{4}$$

which indeed is a U-statistic. Note that under H_0, since $E(U_i) = 0$, it follows that

$E(\frac{X_i}{||X_i||}) = 0$, which implies $E(T_n) = 0$. The above test statistic has an intuitive connection with the work of Bai and Saranadasa (1996) and Chen and Qin (2010), particularly the latter one. To see this, we note that the test statistic of Bai and Saranadasa (1996) for testing (1) is based on $\|\overline{X}\|^2$, while the one of Chen and Qin (2010) is based on $\sum_{i=1}^n \sum_{j=1,j\neq i}^n X_i^T X_j$. By removing the diagonal elements in the statistic of Bai and Saranadasa (1996), Chen and Qin (2010) was able to considerably relax the restrictive condition on p and n. In this spirit, we also dismiss the diagonal elements in defining T_n. Our test statistic hence can be deemed as a nonparametric extension of that of Chen and Qin (2010).

From another point of view, the new test generalizes the multivariate spatial sign test (Brown, 1983; Chaudhuri, 1992; Möttönen and Oja, 1995) for the one-sample location problem to the high-dimensional setting. In the classical setting of $p < n$, Möttönen, Oja and Tienari (1997) derived the asymptotic relative efficiency (ARE) of the spatial sign test versus Hotelling's T^2 test and established its theoretical advantage for heavy-tailed distributions. For example, when the underlying distribution is a 10-dimensional t distribution with ν degrees of freedom, the ARE of the spatial sign test versus the Hotelling's T^2 test is 2.42 when $\nu = 3$, and is 0.95 when $\nu = \infty$ (multivariate normality). However, similarly as Hotelling's T^2 test, the multivariate spatial sign test is not defined when $p > n$. It is an open question whether we can modify it in a way such that its efficiency advantage can be preserved in the high-dimensional setting. This paper provides an affirmative answer.

Remark 1. It is interesting to compare model (2) with the one in Bai and Saranadasa (1996) and Chen and Qin (2010), both of which adopt a factor model structure and a type of pseudo-independence assumption. For example, in Chen and Qin (2010), the factor model assumes that $X = \mu + AZ$, where A is a $p \times m$ matrix for some $m \geq p$ such that $AA^T = \mathrm{Cov}(X)$ and Z satisfies: $E(Z) = 0$ and $Var(Z) = I_m$. Furthermore, for $Z = (Z_1, \ldots, Z_m)^T$, it is assumed that $E(Z_j^4) = 3 + \delta < \infty$, $1 \leq j \leq m$, and that

$$E(Z_{l_1}^{\alpha_1} Z_{l_2}^{\alpha_2} \cdots Z_{l_q}^{\alpha_q}) = E(Z_{l_1}^{\alpha_1}) E(Z_{l_2}^{\alpha_2}) \cdots E(Z_{l_q}^{\alpha_q}) \tag{5}$$

for a positive integer q such that $\sum_{l=1}^{q} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$. Note that although the factor model is flexible, the pseudo-independence assumption (5) is difficult to justify and excludes some commonly-used multivariate distributions such as the multivariate t distribution. In contrast, the models in (2) are rich, based on which one does no need to impose technical condition such as (5) for the asymptotic theory in this paper.

2.1 The limiting null distribution

Despite the simple form of T_n, deriving its asymptotic distribution when $p > n$ is by no means straightforward. As for any other high-dimensional inference, the most challenging issue lies in characterizing the underlying conditions for the asymptotic theory. In Bai and Saranadasa (1996) and Chen and Qin (2010), besides the model structure given in Remark 1, the key sufficient condition is stated through the behavior of the population covariance matrix $\Sigma = \mathrm{Cov}(X)$. In Bai and Saranadasa (1996), it is assumed that $\lambda_{\max}(\Sigma) = o\{\sqrt{\mathrm{Tr}^2(\Sigma^2)}\}$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix and $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. While in Chen and Qin (2010), it is assumed that $\mathrm{Tr}(\Sigma^4) = o\{\mathrm{Tr}^2(\Sigma^2)\}$, which is satisfied under quite relaxed conditions on the eigenvalues of Σ. For the nonparametric test T_n, it is desirable to characterize the underlying conditions in a similar fashion. However, this is very challenging as the building blocks of T_n are the transformations Z_i's, which are not directly related to Σ.

In deriving the asymptotic properties of T_n, moment conditions directly related to Z_i's naturally arise. Lemma 2.1 below plays an important role in this paper. It establishes some of the key properties of the moments of Z_i's under a set of relaxed conditions on the high-dimensional covariance matrix Σ. More specifically, we impose the following two conditions:

(C1) $\mathrm{Tr}(\Sigma^4) = o\{\mathrm{Tr}^2(\Sigma^2)\}$.

(C2) $\frac{\mathrm{Tr}^4(\Sigma)}{\mathrm{Tr}^2(\Sigma^2)} \exp\Big\{ -\frac{\mathrm{Tr}^2(\Sigma)}{128 p \lambda_{\max}^2(\Sigma)} \Big\} = o(1)$.

Remark 2. We first note that conditions (C1) and (C2) both hold trivially if all eigenvalues of Σ are bounded away from 0 and ∞. It is also observed that (C1) holds under some general conditions even if some of the eigenvalues are unbounded (Chen and Qin, 2010). Condition (C2) is quite weak. The exponential term is expected to converge to zero quickly under relaxed conditions. (C2) also permits the eigenvalues to be unbounded. For example, $\lambda_{\max}(\Sigma)$ can diverge to ∞ and the smallest eigenvalue $\lambda_{\min}(\Sigma)$ can converge to zero, both at a polynomial rate of n, provided that $\frac{\mathrm{Tr}(\Sigma)}{\sqrt{p}\lambda_{\max}(\Sigma)}$ diverges to ∞ at a polynomial rate of n.

Lemma 2.1 *Suppose that conditions (C1) and (C2) hold. Let* $B = E\Big(\frac{\epsilon_i \epsilon_i^T}{\|\epsilon_i\|^2}\Big)$. *Then under* H_0*,*

$$E\big\{(Z_1^T Z_2)^4\big\} = o\Big(nE^2\big\{(Z_1^T Z_2)^2\big\}\Big), \tag{6}$$

$$E\big\{(Z_1^T B Z_1)^2\big\} = o\Big(nE^2(Z_1^T B Z_1)\Big), \tag{7}$$

$$E\big\{(Z_1^T B Z_2)^2\big\} = o\Big(E^2\big\{(Z_1^T B Z_1)^2\big\}\Big). \tag{8}$$

The proof of Lemma 2.1 is given in the appendix. The result is established by using a recent probability tool developed by Karoui (2009) on the concentration inequality for quadratic form of the uniform distribution over the unit sphere of $\mathbb{R}^p$.

Some intuition on T_n's asymptotic behavior under H_0 can be gained by observing its first two moments. It is evident that $E(T_n) = 0$. To calculate its variance, we write $T_n = \sum_{i=2}^n Y_i$, where $Y_i = \sum_{j=1}^{i-1} Z_i^T Z_j$. It follows from direct calculation that

$$\begin{aligned} E(Y_i^2) &= \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} E(Z_i^T Z_j Z_i^T Z_k) = \sum_{j=1}^{i-1} E((Z_i^T Z_j)^2) \\ &= (i-1)\mathrm{Tr}(E(Z_1 Z_1^T)E(Z_2 Z_2^T)) = (i-1)\mathrm{Tr}(B^2), \end{aligned}$$

where B is defined in Lemma 2.1. Hence, $\mathrm{Var}(T_n) = \frac{n(n-1)}{2}\mathrm{Tr}(B^2)$. Although T_n has a U-statistics structure, the classical central limit theorem for U-statistics does not apply because the dimension p may depend on the sample size n. By applying Lemma 2.1 and exploring the martingale structure of T_n, we can establish the asymptotic normality of $\frac{T_n}{\sqrt{\mathrm{Var}(T_n)}}$. The limiting null distribution of T_n is given in the following theorem.

Theorem 2.2 *Assume conditions (C1) and(C2) hold. Then under H_0,*

$$\frac{T_n}{\sqrt{\frac{n(n-1)}{2} Tr(B^2)}} \to N(0,1)$$

in distribution, as $n, p \to \infty$.

Remark 3. Note that besides what is implied by (C1) and (C2), the asymptotic normality of T_n under H_0 holds without explicit conditions on the relationship between p and n. (C1) and (C2) both hold as long as the eigenvalues of Σ are bounded away from 0 and ∞. They are generally weaker than those conditions in the literature which explicitly imposed a relationship between n and p such as $p = o(n^2)$. Furthermore, the above asymptotic normality result can be extended beyond the family of elliptical distributions. In (2), the requirement that U_i is uniformly distributed on the sphere can be relaxed. In fact, concentration inequalities similar to that given in Lemma A.2, which is the key to the proof, can be obtained by random vectors that satisfy certain concentration of measure properties (Karoui, 2009).

Remark 4. To apply T_n in practical data analysis, we need an estimator of $\mathrm{Tr}(B^2)$. Following Chen and Qin (2010), we may estimate $\mathrm{Tr}(B^2)$ using the cross-validation approach as follows.

$$\widehat{\mathrm{Tr}(B^2)} = \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\sum_{1\le j\ne k\le n}(Z_j - \overline{Z}_{(j,k)})Z_j^T(Z_j - \overline{Z}_{(j,k)})Z_k^T\Big\}, \tag{9}$$

where $\overline{Z}_{(j,k)}$ is the sample mean after excluding Z_j and Z_k. It is noteworthy that the estimator in Chen and Qin can be computationally intensive for large p as each term inside the U-statistic involves multiplying high-dimensional matrices. In contrast, the computational burden of the estimator in (9) can be substantially reduced by observing that $\|Z_j\|^2 = 1$. Let $\overline{Z}^* = (n-2)^{-1}\sum_{m=1}^n Z_m$. In the Appendix, it is derived that

$$\begin{aligned}\widehat{\mathrm{Tr}(B^2)} \;=\;& -\frac{n}{(n-2)^2} + \frac{(n-1)}{n(n-2)^2}\mathrm{Tr}\{(\sum_{j=1}^n Z_jZ_j^T)^2\} + \frac{1-2n}{n(n-1)}\overline{Z}^{*T}(\sum_{j=1}^n Z_jZ_j^T)\overline{Z}^* \\ &+\frac{2}{n}\left\|\overline{Z}^*\right\|^2 + \frac{(n-2)^2}{n(n-1)}\left\|\overline{Z}^*\right\|^4. \end{aligned}\tag{10}$$

2.2 Local power comparison

We now turn our attention to the power analysis of T_n under contiguous sequences of alternative hypotheses. This analysis enables us to further investigate the asymptotic relative efficiency of T_n with respect to Chen and Qin's test (referred to as CQ test in the sequel). Some interesting findings are revealed, which suggests promising efficiency gain of the new test for heavy-tailed multivariate distributions.

For the local power analysis, we impose the following additional conditions.

(C3) $\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{256p\lambda^2_{\max}(\Sigma)}\Big) = o\Big(\min\Big(\frac{\lambda_{\max}(\Sigma)}{\mathrm{Tr}(\Sigma)}, \frac{\lambda_{\min}(\Sigma)}{\lambda_{\max}(\Sigma)}\Big)\Big)$.

(C4) $\lambda_{\max}(\Sigma) = o(\mathrm{Tr}(\Sigma))$.

(C5) $\|\mu\|^2 E(\|\epsilon\|^{-2}) = o\big(\min\big(n^{-1}\frac{\mathrm{Tr}^2(\Sigma)}{\lambda_{\max}(\Sigma)\mathrm{Tr}(\Sigma)}, n^{-1/2}\frac{\mathrm{Tr}^{1/2}(\Sigma^2)}{\mathrm{Tr}(\Sigma)}\big)\big)$.

(C6) For some $0 < \delta < 1$, $\|\mu\|^{2\delta} E(\|\epsilon\|^{-2-2\delta}) = o\big(E^2(\|\epsilon\|^{-1})\big)$.

Remark. Conditions (C3) and (C4) are concerned with the properties of the population covariance matrix Σ. These two conditions are relatively weak. In particular, they are satisfied when the eigenvalues of Σ are bounded away from 0 and ∞. Conditions (C5) and (C6) can be viewed as high-dimensional local-alternative statements for $p > n$. To gain some insight into the local alternative, we consider the case the eigenvalues of Σ are bounded away from 0 and ∞, then the right-hand side of (C5) is $o\big(n^{-1/2}p^{-1/2}\big)$. For p-dimensional spherical t-distribution with ν degrees of freedom, $\frac{p}{\|\epsilon\|^2} \sim F(\nu, p)$. It is easy to show that $E(\|\epsilon\|^{-2}) = \frac{1}{p-2}$. A slightly more involved calculation based on the properties of F-distribution reveals that $E(\|\epsilon\|^{-1}) = O(p^{-1/2})$ and $E(\|\epsilon\|^{-2-2\delta}) = O(p^{-1-\delta})$. Then the conditions in (C4) and (C5) amount to $\|\mu\|^2 = o(n^{-1/2}p^{1/2})$ and $\|\mu\|^{2\delta} = o(p^{\delta})$ for some $0 < \delta < 1$. If $\delta = 1/2$, then the condition further reduces to $\|\mu\| = o(n^{-1/4}p^{1/4})$. If we consider the local alternatives such that all components of μ are equal to κ, then we have $\kappa = o(n^{-1/4}p^{-1/4})$, which when $p > n$ is of smaller order of $n^{-1/2}$, the usual local alternative rate for Hotelling's test with fixed dimension. The faster rate of local alternative can be viewed as a blessing of high dimensionality, where more information can be gained to distinguish subtle deviation from the null hypothesis.

Theorem 2.3 *Assume conditions (C1)-(C6) hold. Letting* $A = E\Big\{\frac{1}{\|\epsilon_i\|}\big(I_p - \frac{\epsilon_i\epsilon_i^T}{\|\epsilon_i\|^2}\big)\Big\}$. *Then as* $n, p \to \infty$,

$$\frac{T_n - \frac{n(n-1)}{2}\mu^T A^2 \mu(1+o(1))}{\sqrt{\frac{n(n-1)}{2} Tr(B^2)}} \to N(0,1)$$

in distribution.

Theorem 2.3 implies that under the local alternatives, the proposed level α test has the local power

$$\beta_n = \Phi\Big(- z_\alpha + \sqrt{\frac{n(n-1)}{2}} \frac{\mu^T A^2 \mu(1+o(1))}{\sqrt{\mathrm{Tr}(B^2)}}\Big),$$

where $\Phi(\cdot)$ and z_α denote the cumulative distribution function and the upper α quantile of the $N(0,1)$ distribution, respectively. On the other hand, the test of CQ test has the local power

$$\beta_n^{\mathrm{CQ}} = \Phi\Big(- z_\alpha + \frac{n\,\|\mu\|^2}{\sqrt{2\mathrm{Tr}(\Sigma^2)}}\Big). \tag{11}$$

The asymptotic relative efficiency (ARE) of T_n versus the CQ test is

$$\mathrm{ARE}_{T_n,\mathrm{CQ}} = \frac{\mu^T A^2 \mu}{\|\mu\|^2} \sqrt{\frac{\mathrm{Tr}(\Sigma^2)}{\mathrm{Tr}(B^2)}}(1+o(1)). \tag{12}$$

To give an idea of the implication of the above result, we consider the asymptotic relative efficiency when the data arise from a spherical p-dimensional t distribution with ν degrees of freedom ($\nu > 2$). In this case, $A = E[\|\epsilon\|^{-1}]\frac{p-1}{p} I_p$ where I_p denotes the $p \times p$ identity matrix, $\mathrm{Tr}(B^2) = p^{-1}$ and $\mathrm{Tr}(\Sigma^2) = p^{-1}E^2[\|\epsilon\|^2]$. Hence, (12) reduces to

$$\mathrm{ARE}_{T_n,\mathrm{CQ}} = p^{-2}(p-1)^2 E^2[\|\epsilon\|^{-1}]E[\|\epsilon\|^2] \tag{13}$$

For the t distribution, we have $E[\|\epsilon\|^2] = \frac{p\nu}{\nu-2}$ and $E[\|\epsilon\|^{-1}] = \sqrt{\frac{2}{p\nu}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}$, where $\Gamma(t) = \int_0^\infty u^{t-1}e^{-u}du$ denotes the gamma function. For large p, the asymptotic relative efficiency thus is approximately

$$\mathrm{ARE}_{T_n,\mathrm{CQ}} \approx \frac{2}{\nu-2}\Big(\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\Big)^2.$$

For $\nu = 3$, this value is about 2.54; for $\nu = 4$, it is about 1.76; for $\nu = 5$, it is about 1.51; for $\nu = 6$, it is about 1.38; for $\nu = \infty$ (corresponding to multivariate normal distribution), by noting that $\Gamma((\nu+1)/2) \approx \Gamma(\nu/2)\sqrt{\frac{\nu}{2}}$ as $\nu \to \infty$, we have that the ARE has limit one. Theoretically, the efficiency loss of the new test under multivariate normality is little, but the efficiency gain can be substantial for heavy-tailed distribution.

3 Numerical studies

3.1 Monte Carlo simulations

In this section, we compare the performance of the new test with the test of Chen and Qin (CQ test, 2010), the test based on multiple comparison with Bonferroni correction which controls the family error rate at 0.05, and the test based on multiple comparison with FDR control (Benjamini and Hochberg, 1995). These three tests are labeled as CQ, BF and FDR respectively in Tables 1, 2 and 3. The performance of the four tests are evaluated on 1000 simulation runs. We consider $n = 20, 50$ and $p = 200, 1000$ and 2000.

Example 1. In this example, data were generated from $N_p(\mu, \Sigma)$, where $\Sigma = (\sigma_{ij})$ has the structure $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.2$ $(i \neq j)$. We consider the following three cases:

Case 1: data are generated under the null hypothesis $\mu = (0, \ldots, 0)^T$;

Case 2: data are generated under the alternative $\mu = (0.25, 0.25, \ldots, 0.25)^T$;

Case 3: data are generated under the alternative $\mu = (\mu_1, \ldots, \mu_p)^T$ has the form $\mu_1 = \ldots = \mu_{\frac{p}{3}} = 0$, $\mu_{\frac{p}{3}+1} = \ldots = \mu_{\frac{2p}{3}} = 0.2$ and $\mu_{\frac{2p}{3}+1} = \ldots = \mu_p = -0.2$.

Note that the Monte Carlo error is 0.0135 for 1000 simulation at level 0.05. Table 1 indicates that the four tests have nominal levels reasonably close to 0.05, especially when $n = 50$. For the alternative in Case 2, the performance of the new test is very close to that of the CQ test, both are significantly better than the multiple comparison based Bonferroni test and FDR test. The latter two test have especially low power when $n = 20$. For the alternative in Case 3, all four test have fine performance when $n = 50$; while when $n = 20$, the new test has slightly higher power than the CQ test, both are substantially better than the Bonferroni test and FDR test.

Example 2. Similar as Example 1 but with X_j from a p-variate t distribution with mean vector μ, covariance matrix Σ and 3 degrees of freedom, where μ and Σ are the same as the ones in Example 1. In this example, the distribution is heavy-tailed. Both the new test and the CQ test perform reasonably well under the null hypothesis, while the Bonferroni test and FDR test are both somewhat conservative. In this example, the Bonferroni test and FDR test perform poorly under the two alternatively. The new test displays substantially higher power than that of the CQ test for both alternatives. It is noted that for the second alternative, the power of the new test is more than twice of that of the CQ test when $n = 20$.

Table 1: Simulation results for normal distribution

Case	n	p	New	CQ	BF	FDR
1	20	200	0.064	0.071	0.029	0.034
	20	1000	0.066	0.069	0.046	0.046
	20	2000	0.073	0.070	0.052	0.053
	50	200	0.049	0.057	0.043	0.043
	50	1000	0.059	0.060	0.035	0.038
	50	2000	0.058	0.061	0.043	0.047
2	20	200	0.708	0.728	0.353	0.403
	20	1000	0.723	0.723	0.405	0.471
	20	2000	0.720	0.729	0.385	0.447
	50	200	0.965	0.970	0.818	0.861
	50	1000	0.975	0.976	0.842	0.890
	50	2000	0.970	0.976	0.850	0.901
3	20	200	0.908	0.768	0.312	0.357
	20	1000	0.951	0.826	0.382	0.443
	20	2000	0.954	0.821	0.404	0.464
	50	200	1.000	1.000	0.913	0.975
	50	1000	1.000	1.000	0.964	0.997
	50	2000	1.000	1.000	0.973	0.998

Example 3. In this example, data were generated from a scaled mixture of normal distributions $0.9 * N_p(\mu, \Sigma) + 0.1 * N_p(\mu, 3^2\Sigma)$, where μ and Σ are the same as the ones in Example 1. The distribution in this example also has heavy tails. Similarly as in Example 2, the new test significantly outperforms the three contending approaches.

3.2 An application

Type 2 diabetes is one of the most common chronic diseases. Insulin resistance in skeletal muscle, which is the major site of glucose disposal, is a prominent feature of Type 2 diabetes. To study insulins ability to regulate gene expression, an experiment performed microarray analysis using the Affymetrix Hu95A chip of human skeletal muscle biopsies from 15 diabetic patients both before and after insulin treatment (Wu et al., 2007). The gene expression alterations are promising to provide insights on

Table 2: Simulation results for multivariate t-distribution

Case	n	p	New	CQ	BF	FDR
1	20	200	0.073	0.078	0.017	0.020
	20	1000	0.083	0.088	0.011	0.013
	20	2000	0.064	0.072	0.010	0.010
	50	200	0.054	0.050	0.024	0.026
	50	1000	0.053	0.063	0.011	0.012
	50	2000	0.069	0.076	0.010	0.010
2	20	200	0.579	0.419	0.138	0.162
	20	1000	0.633	0.472	0.153	0.183
	20	2000	0.631	0.468	0.117	0.138
	50	200	0.937	0.713	0.419	0.477
	50	1000	0.941	0.721	0.424	0.491
	50	2000	0.921	0.736	0.438	0.492
3	20	200	0.758	0.304	0.110	0.118
	20	1000	0.815	0.371	0.129	0.15
	20	2000	0.830	0.363	0.107	0.122
	50	200	1.000	0.775	0.407	0.460
	50	1000	1.000	0.803	0.427	0.485
	50	2000	1.000	0.825	0.493	0.571

new therapeutic targets for the treatment of this common disease. Hence, we are interesting in testing the hypothesis in (1), where μ represents the average change of the gene expression level due to the treatment.

The underlying genetics of Type 2 diabetes were recognized to be very complex. It is believed that Type 2 diabetes is resulted from interactions between many genetic factors and the environment. In our analysis, we considered 2519 curated gene sets. The gene sets we used are from the C2 collection of the GSEA online pathway databases (http://www.broadinstitute.org/gsea/msigdb/collection_details.jsp#C2). The largest gene set contains 1607 genes, which makes the hypothesis testing problem a high-dimensional one.

We applied both the new test and the CQ test at 5% significance level with the Bonferroni correction to control the family-wise error rate at 0.05 level. For the CQ

Table 3: Simulation for scaled mixture of normals

Case	n	p	New	CQ	BF	FDR
1	20	200	0.070	0.066	0.023	0.024
	20	1000	0.063	0.070	0.014	0.015
	20	2000	0.045	0.049	0.015	0.018
	50	200	0.047	0.049	0.017	0.022
	50	1000	0.063	0.066	0.02	0.021
	50	2000	0.042	0.040	0.015	0.016
2	20	200	0.618	0.530	0.218	0.250
	20	1000	0.649	0.548	0.209	0.259
	20	2000	0.627	0.542	0.193	0.231
	50	200	0.942	0.830	0.568	0.633
	50	1000	0.941	0.859	0.600	0.682
	50	2000	0.964	0.867	0.619	0.687
3	20	200	0.789	0.446	0.184	0.200
	20	1000	0.870	0.449	0.175	0.201
	20	2000	0.882	0.492	0.224	0.243
	50	200	1.000	0.944	0.608	0.689
	50	1000	1.000	0.966	0.663	0.759
	50	2000	1.000	0.968	0.698	0.800

Table 4: Top 10 significant gene sets selected by the new test and CQ test

Gene sets	New test	CQ test
ZWANG_CLASS_2_TRANSIENTLY_INDUCED_BY_EGF	24.34	20.11
NAGASHIMA_EGF_SIGNALING_UP	22.44	17.01
SHIPP_DLBCL_CURED_VS_FATAL_DN	22.34	18.24
WILLERT_WNT_SIGNALING	19.66	NA
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	19.63	18.46
PID_HIF2PATHWAY	19.46	15.65
PHONG_TNF_TARGETS_UP	19.21	18.90
AMIT_EGF_RESPONSE_60_HELA	18.64	16.38
MCCLUNG_CREB1_TARGETS_DN	18.43	NA
SEMENZA_HIF1_TARGETS	18.34	NA
AMIT_SERUM_RESPONSE_40_MCF10A	NA	15.98
AMIT_SERUM_RESPONSE_60_MCF10A	NA	15.43
PLASARI_TGFB1_TARGETS_1HR_UP	NA	15.00

method, 520 gene sets (20.64% of all candidates) are identified as significance; and for the new method, 954 gene sets (37.87% of all candidates) are selected as significant. We observe that the significant gene sets selected by the new test include those identified by

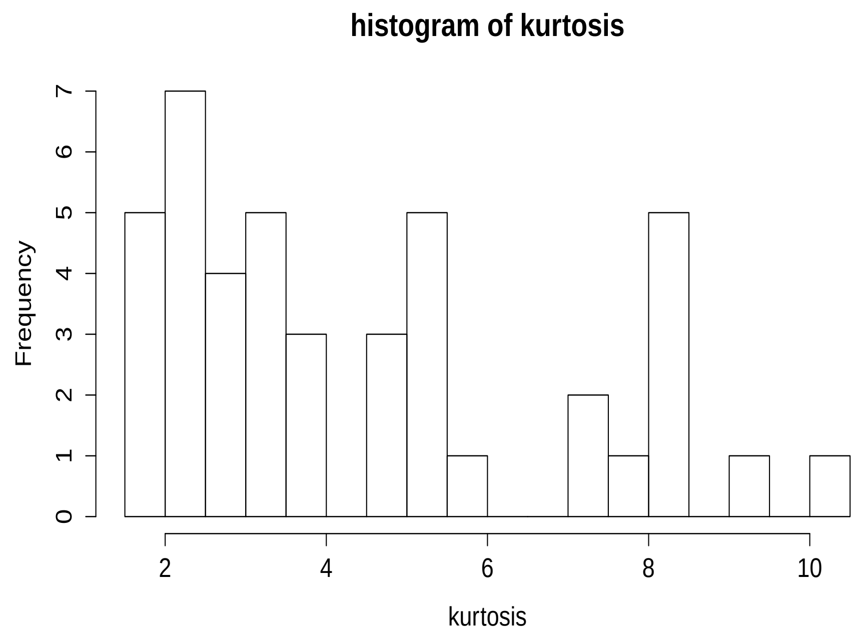

Figure 1: The histogram of marginal kurtosises for all genes in MCCLUNG_CREB1_TARGETS_DN gene set.

the CQ test with only one exception (HASLINGER_B_CLL_WITH_CHROMOSOME _12_TRISOMY).

Table 3 displays the top 10 significant gene sets identified by the two tests and their corresponding test statistics values. We observe these two lists share 7 common gene sets. Among these seven gene sets, ZWANG_CLASS_2_TRANSIENTLY_INDUCED _BY_EGF, NAGASHIMA_EGF_SIGNALING_UP, AMIT_EGF_RESPONSE _60_HELA, AMIT_SERUM_RESPONSE_40_MCF10A and AMIT_SER UM_RESPONSE_60_MCF10A are known to be biologically related to insulin effect on human cells. The gene set SEMENZA_HIF1_TARGETS is only on the top ten list of the new test and was also found to be biologically related to insulin effect on human cells. Most of those significant gene sets are induced by Epidermal growth factor (EGF) or insulin-like growth factor (IGF).

It is interesting to point out that exploratory analysis of the gene expression data suggests the multivariate normality assumption is questionable. For example, Figure 1 displays the histogram of the marginal kurtosises of the difference of each gene expression levels (before/after the treatment) of all genes in MCCLUNG_CREB1_TARGETS_DN gene set, which was selected among the top 10 gene sets by the new method but not by the CQ method. Figure 1 clearly shows that some gene expression levels have heavy tails as their kurtosises are much larger than 3, the kurtosis of a normal distribution.

References

[1] Bai, Z., and Sarandasa, H. (1996), "Effect of High Dimension: By an Example of a Two Sample Problem". *Statistica Sinica, 6, 311-329.*

[2] Brown, B. M. (1983), "Statistical Uses of the Spatial Median," *Journal of the Royal Statistical Society, Series B*, 45, 25-30.

[3] Chaudhuri, P. (1992), "Multivariate Location Estimation Using Extension of R-estimates through U-statistics Type Approach," *Annals of Statistics*, 20, 897-916.

[4] Chen, S. X., and Qin, Y. L. (2010), "A Two-sample Test for High-dimensional Data with Application to Gene-Set Testing," *Annals of Statistics*, 38, 808-835.

[5] Fang, K. T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

[6] Hall, P., and Heyde, C. (1980), *Martingale Limit Theory and Applications*, Academic Press, New York.

[7] Karoui, E. Noureddine. (2009), "Concentration of Measure and Spectra of Random Matrices: with Applications to Correlation Matrices, Elliptical Distributions and Beyond," *The Annals of Applied Probability*, 19, 2362-2405.

[8] Ledoux, M. (2001). *The Concentration of Measure Phenomenon.* American Mathematical Society. Providence, Rhode Island.

[9] Mcneil, A. J., Frey, R., and Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, NJ.

[10] Möttönen, J., and Oja, H. (1995), "Multivariate Spatial Sign and Rank Methods," *Journal of Nonparametric Statistics*, 5, 201-213.

[11] Möttönen, J., Oja, H., and Tienari, J. (1997), "On the Efficiency of Multivariate Spatial Sign and Rank Tests," *Annals of Statistics*, 25, 542-552.

[12] Oja, H. (2010), *Multivariate Nonparametric Methods with R*, Springer.

[13] Schmidt, R. (2002), "Tail Dependence for Elliptically Contoured Distributions," *Mathematical Methods of Operations Research*, 55, 301-327.

[14] Srivastava, M. S., and Du, M. (2008), "A Test for the Mean Vector with Fewer Observations than the Dimension," *Journal of Multivariate Analysis*, 99, 386-402.

[15] Wu, X., Wang, J., Cui, X., Maianu, L. et al., (2007), "The Effect of Insulin on Expression of Genes and Biochemical Pathways in Human Skeletal Muscle," *Endocrine*, 31, 5-17.

Appendix: Technical proofs

Appendix 1: Some useful lemmas

Lemma A.1 *Let $U = (U_1, \dots, U_p)^T$ be a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$. Then*

(1) $E(U) = 0$, $Var(U) = p^{-1}I_p$, $E(U_j^4) = \frac{3}{p(p+2)}$, $\forall\ j$, and $E(U_j^2U_k^2) = \frac{1}{p(p+2)}$ for $j \neq k$.

(2) Let M be a deterministic real-valued matrix. Assume that $\|M\|_2 \leq k$, where $\|M\|_2$ denotes the spectral norm of M. Then, $\forall\ t > 0$, $P\Big(\big|U^TMU - p^{-1}Tr(M)\big| > t\Big) \leq 2\exp\Big(-\frac{(p-1)(t-c_p)^2}{8k^2}\Big)$, where $c_p = \sqrt{\frac{8\pi k^2}{p-1}}$.

Proof. Part (1) was shown in Section 3.1 of Fang, Kotz and Ng (1990). Part (2) was implied by Lemma 7 of Karoui (2009), which deals with a more general setting and a complex-valued matrix M. For completeness, we include below a simple, direct proof based on Karoui's idea. First, note that $\big|U_1^TMU_1 - U_2^TMU_2\big| = \big|U_1^TM(U_1-U_2)+(U_1-U_2)^TMU_2\big| \leq \|U_1-U_2\|_2\|M\|_2(\|U_1\|_2+\|U_2\|_2) \leq 2k\|U_1-U_2\|_2$. Hence, U^TMU is a $2k$-Lipschitz function of U on the unit sphere. Theorem 2.3 of Ledoux (2001) implies that $P\Big(\big|U^TMU - \text{median}(U^TMU)\big| > t\Big) \leq 2\exp\big(-\frac{(p-1)t^2}{8k^2}\big)$, $\forall\ t > 0$. Furthermore, Proposition 1.9 of Ledoux (2001) implies $\big|\text{median}(U^TMU) - p^{-1}\text{Tr}(M)\big| \leq \sqrt{\frac{8\pi k^2}{(p-1)}}$. These two inequalities together give the result in part (2). $\square$

Lemma A.2 *(A concentration inequality) Assume $W = \Gamma U$, where U is uniformly distributed on the unit sphere in $\mathbb{R}^p$. Let $\Omega = \Gamma\Gamma^T$ and consider the event $A = \{\frac{Tr(\Omega)}{2p} \le \|W\|^2 \le \frac{3\,Tr(\Omega)}{2p}\}$. Then*

$$P(A) \ge 1 - c_1 \exp\Big(-\frac{Tr^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big), \qquad \text{(A.1)}$$

for all $p > 1$, where $c_1 = 2\exp(\pi/2)$ is a finite constant.

Proof. We have $\|W\|^2 = U^T M U$ where $M = \Gamma^T\Gamma$. Note that $\|M\|_2 = \lambda_{\max}^{1/2}((\Gamma^T\Gamma)^2) = \lambda_{\max}(\Omega)$. Since $A = \big\{\big|\,\|W\|^2 - p^{-1}\mathrm{Tr}(\Omega)\big| \le (2p)^{-1}\mathrm{Tr}(\Omega)\big\}$, by taking $t = (2p)^{-1}\mathrm{Tr}(\Omega)$ in part (2) of Lemma A.1, we have

$$P(A) \ge 1 - 2\exp\Big(-k^{-2}(p-1)\big(\frac{\mathrm{Tr}(\Omega)}{2p} - c_p\big)^2/16\Big), \qquad \text{(A.2)}$$

where $c_p = \sqrt{\frac{8\pi k^2}{p-1}}$ and $k = \lambda_{\max}(\Omega)$. Using the fact $(a-b)^2 \ge \frac{1}{2}a^2 - b^2$, for any $a > 0$, $b > 0$, we have

$$\begin{aligned}
&\exp\Big(-k^{-2}(p-1)\big(\frac{\mathrm{Tr}(\Omega)}{2p} - c_p\big)^2/8\Big) = \exp\Big(-(p-1)\Big(\frac{\mathrm{Tr}(\Omega)}{2p\lambda_{\max}(\Omega)} - \sqrt{\frac{8\pi}{p-1}}\Big)^2/8\Big)\\
\le\ &\exp\Big(-(p-1)\Big(\frac{\mathrm{Tr}^2(\Omega)}{8p^2\lambda_{\max}^2(\Omega)} - \frac{8\pi}{p-1}\Big)/8\Big)\\
\le\ &\exp(\pi/2)\exp\Big(-\frac{\mathrm{Tr}^2(\Omega)}{128p\lambda_{\max}^2(\Omega)}\Big) = \exp\big(\frac{\pi}{2}\big)\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big)
\end{aligned}$$

for all $p > 1$. Hence the result follows. □

Proof of Lemma 2.1. Under H_0, $Z_i = W_i/\|W_i\|$, $i = 1, 2$, where $W_i = \Gamma U_i$, and U_1 and U_2 are independent random vectors uniformly distributed on the unit sphere in $\mathbb{R}^p$. We have $E(\|W_i\|^2) = E(U_i^T\Gamma^T\Gamma U_i) = \mathrm{Tr}(\Gamma^T\Gamma E(U_iU_i^T)) = p^{-1}\mathrm{Tr}(\Omega)$, $i = 1, 2$, by part (1) of Lemma A.1.

We first prove (6). Note that $E\{(Z_1^TZ_2)^4\} = E\Big\{\frac{(W_1^TW_2)^4}{\|W_1\|^4\|W_2\|^4}\Big\}$. To evaluate $E\{(W_1^TW_2)^4\}$, we write $\Gamma^T\Gamma = (\nu_{ij})_{p\times p}$. Then

$$
\begin{aligned}
& E\big\{(W_1^T W_2)^4\big\} = E\Big\{\big(\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}U_{1i}U_{2j}\big)^4\Big\} \\
= \ & \sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4 E^2(U_{1i}^4) + 3 \sum_{\substack{1\le i\le p\\ 1\le j_1\ne j_2\le p}} \nu_{ij_1}^2\nu_{ij_2}^2 E(U_{1i}^4)E(U_{2j_1}^2U_{2j_2}^2) \\
& +3 \sum_{\substack{1\le j\le p\\ 1\le i_1\ne i_2\le p}} \nu_{i_1j}^2\nu_{i_2j}^2 E(U_{2j}^4)E(U_{1i_1}^2U_{1i_2}^2) \\
& +O(1) \sum_{1\le i_1\ne i_2\le p}\ \sum_{1\le j_1\ne j_2\le p} (\nu_{i_1j_1}^2\nu_{i_2j_2}^2 + \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2})E(U_{1i_1}^2U_{1i_2}^2)E(U_{2j_1}^2U_{2j_2}^2) \\
= \ & O(p^{-4})\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4 + O(p^{-4}) \sum_{\substack{1\le i\le p\\ 1\le j_1\ne j_2\le p}} \nu_{ij_1}^2\nu_{ij_2}^2 \\
& +O(p^{-4}) \sum_{1\le i_1\ne i_2\le p}\ \sum_{1\le j_1\ne j_2\le p} (\nu_{i_1j_1}^2\nu_{i_2j_2}^2 + \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2}) \\
\le \ & O(p^{-4})\mathrm{Tr}^2(\Omega^2) + O(p^{-4})\mathrm{Tr}(\Omega^4) \le O(p^{-4})\mathrm{Tr}^2(\Omega^2)
\end{aligned}
$$

by part (1) of Lemma A.1 and condition (C1), and noticing that

$$
\begin{aligned}
& \max\Big\{\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4,\ \sum_{\substack{1\le i\le p\\ 1\le j_1\ne j_2\le p}} \nu_{ij_1}^2\nu_{ij_2}^2,\ \sum_{1\le i_1\ne i_2\le p}\ \sum_{1\le j_1\ne j_2\le p} \nu_{i_1j_1}^2\nu_{i_2j_2}^2\Big\} \\
\le \ & \big(\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^2\big)^2 = \big(\mathrm{Tr}(\Gamma^T\Gamma\Gamma^T\Gamma)\big)^2 = \mathrm{Tr}^2(\Omega^2)
\end{aligned}
$$

and

$$
\begin{aligned}
& \sum_{1\le i_1\ne i_2\le p}\ \sum_{1\le j_1\ne j_2\le p} \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2} \\
\le \ & \sum_{1\le i_1\ne i_2\le p} \nu_{i_1i_2}^{(2)}\nu_{i_1i_2}^{(2)} \le \sum_{i=1}^p \nu_{i_1i_1}^{(4)} = \mathrm{Tr}(\Gamma^T\Omega\Gamma\Gamma^T\Omega\Gamma) = \mathrm{Tr}(\Omega^4),
\end{aligned}
$$

where we write $\Gamma^T\Omega\Gamma = (\nu_{ij}^{(2)})_{p\times p}$ and $(\Gamma^T\Omega\Gamma)^2 = (\nu_{ij}^{(4)})_{p\times p}$.

Let $A_i = \big\{\frac{\mathrm{Tr}(\Omega)}{2p} \le \|W_i\|^2 \le \frac{3\mathrm{Tr}(\Omega)}{2p}\big\}$, $i = 1, 2$. Applying the above upper bound of

$E\{(W_1^TW_2)^4\}$, we have

$$
\begin{aligned}
E\Big\{\frac{(W_1^TW_2)^4}{\|W_1\|^4\,\|W_2\|^4}\Big\} &= E\Big\{\frac{(W_1^TW_2)^4}{\|W_1\|^4\,\|W_2\|^4}I(A_1\cap A_2)\Big\} + E\Big\{\frac{(W_1^TW_2)^4}{\|W_1\|^4\,\|W_2\|^4}I(A_1^c\cup A_2^c)\Big\} \\
&\le \frac{E\{(W_1^TW_2)^4\}}{(16p^4)^{-1}\mathrm{Tr}^4(\Omega)} + P(A_1^c\cup A_2^c) \\
&\le \frac{O(p^{-4}\mathrm{Tr}^2(\Omega^2))}{(16p^4)^{-1}\mathrm{Tr}^4(\Omega)} + 2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda^2_{\max}(\Sigma)}\Big),
\end{aligned}
$$

where c_1 is the constant in Lemma A.2, the first inequality follows from the Cauchy-Schwarz inequality and the second inequality is a result of Lemma A.2.

On the other hand, $E^2\{(Z_1^TZ_2)^2\} = E^2\Big\{\frac{(W_1^TW_2)^2}{\|W_1\|^2\|W_2\|^2}\Big\}$. Write $\Gamma = (\Gamma_1,\ldots,\Gamma_p)^T$, that is, Γ_j^T denotes the jth row of Γ, $j=1,\ldots,p$. Then it can be shown that

$$
E\{(W_1^TW_2)^2\} = p^{-2}\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}(\Gamma_{k_1}^T\Gamma_{k_2})^2 = p^{-2}\mathrm{Tr}(\Omega^2).
$$

Furthermore, applying Lemma A.2, we have

$$
\begin{aligned}
& E\Big\{\frac{(W_1^TW_2)^2}{\|W_1\|^2\,\|W_2\|^2}\Big\} \ge E\Big\{\frac{(W_1^TW_2)^2}{\|W_1\|^2\,\|W_2\|^2}I(A_1\cap A_2)\Big\} \\
\ge\; & \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}E\Big\{(W_1^TW_2)^2I(A_1\cap A_2)\Big\} \\
=\; & \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}E\Big\{(W_1^TW_2)^2\Big\} - \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}E\Big\{(W_1^TW_2)^2I(A_1^c\cup A_2^c)\Big\} \\
\ge\; & \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}E\Big\{(W_1^TW_2)^2\Big\} - \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}E^{1/2}\Big\{(W_1^TW_2)^4\Big\}E^{1/2}\Big\{I(A_1^c\cup A_2^c)\Big\} \\
\ge\; & \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}\frac{\mathrm{Tr}(\Omega^2)}{p^2} - \frac{4p^2/9}{\mathrm{Tr}^2(\Omega)}O(p^{-2})\mathrm{Tr}(\Omega^2)\sqrt{2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda^2_{\max}(\Sigma)}\Big)} \\
=\; & \frac{4\mathrm{Tr}(\Omega^2)}{9\mathrm{Tr}^2(\Omega)}(1-o(1)).
\end{aligned}
$$

In the above, the third inequality applies the Hölder's inequality. To prove the last equality, we use condition (C2) by noting that the fact $\mathrm{Tr}(\Sigma^2)\le\mathrm{Tr}^2(\Sigma)$ implies

$\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{256p\lambda^2_{\max}(\Sigma)}\Big)=o(1)$. Hence,

$$
\begin{aligned}
\frac{E\big\{(Z_1^TZ_2)^4\big\}}{nE^2\big\{(Z_1^TZ_2)^2\big\}} &= \frac{E\Big\{\frac{(W_1^TW_2)^4}{\|W_1\|^4\|W_2\|^4}\Big\}}{nE^2\Big\{\frac{(W_1^TW_2)^2}{\|W_1\|^2\|W_2\|^2}\Big\}} \le \frac{\frac{O(p^{-4}\mathrm{Tr}^2(\Omega^2))}{(16p^4)^{-1}\mathrm{Tr}^4(\Omega)}+2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{256p\lambda^2_{\max}(\Sigma)}\Big)}{n(\frac{4\mathrm{Tr}(\Omega^2)}{9\mathrm{Tr}^2(\Omega)})^2(1-o(1))} \\
\le\ & O(n^{-1})+O(n^{-1})\frac{\mathrm{Tr}^4(\Sigma)}{\mathrm{Tr}^2(\Sigma^2)}2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda^2_{\max}(\Sigma)}\Big)=o(1),
\end{aligned}
$$

by condition (C2). This proves (6).

Next we prove (7). Note that $E\big(Z_1^TBZ_1\big)=E\big\{(Z_1^TZ_2)^2\big\}$ and

$$
\begin{aligned}
& E\big\{(Z_1^TBZ_1)^2\big\}=E\big\{(Z_1^TE(Z_2Z_2^T)Z_1)^2\big\}=E\big\{E^2(Z_1^TZ_2Z_2^TZ_1|Z_1)\big\} \\
\le\ & E\big[E\{(Z_1^TZ_2Z_2^TZ_1)^2|Z_1)\}\big]=E\{(Z_1^TZ_2Z_2^TZ_1)^2\}=E\{(Z_1^TZ_2)^4\}.
\end{aligned}
$$

Hence, (7) follows from (6).

Finally we prove (8). Note that

$$
\begin{aligned}
E\big\{(Z_1^TBZ_2)^2\big\} &= E\big\{(Z_1^TE(Z_3Z_3^T)Z_2)^2\big\}=E\big\{E^2(Z_1^TZ_3Z_3^TZ_2|Z_1,Z_2)\big\} \\
&= E\Big\{E^2\Big(\frac{W_1^TW_3W_3^TW_2}{\|W_1\|\cdot\|W_2\|\cdot\|W_3\|^2}\Big|W_1,W_2\Big)\Big\}.
\end{aligned}
$$

Similarly as for the proof of (6), we can show that

$$
\frac{E\big\{(Z_1^TBZ_2)^2\big\}}{E^2\big\{(Z_1^TBZ_2)\big\}} \le \frac{\frac{p^{-2}E\big\{(W_1^T\Omega W_2)^2\big\}}{(16p^4)^{-1}\mathrm{Tr}^4(\Omega)}+6c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda^2_{\max}(\Sigma)}\Big)}{(\frac{4\mathrm{Tr}(\Omega^2)}{9\mathrm{Tr}^2(\Omega)})^2(1-o(1))}.
$$

Write $\Gamma^T\Omega\Gamma=\sum_{l=1}^p\xi_l\xi_l^T$, then

$$
\begin{aligned}
& E\big\{(W_1^T\Omega W_2)^2\big\}=E\big\{(U_1^T\Gamma^T\Omega\Gamma U_2)^2\big\}=\sum_{l_1=1}^p\sum_{l_2=1}^pE\big(U_1^T\xi_{l_1}\xi_{l_1}^TU_2U_1^T\xi_{l_2}\xi_{l_2}^TU_2\big) \\
=\ & \sum_{l_1=1}^p\sum_{l_2=1}^pE\big(U_1^T\xi_{l_1}\xi_{l_2}^TU_1\big)E\big(U_2^T\xi_{l_1}\xi_{l_2}^TU_2\big)=p^{-2}\sum_{l_1=1}^p\sum_{l_2=1}^p(\xi_{l_1}^T\xi_{l_2})^2=p^{-2}\mathrm{Tr}(\Omega^4).
\end{aligned}
$$

Hence, by conditions (C1) and (C2),

$$\frac{E\big\{(Z_1^T B Z_2)^2\big\}}{E^2\big\{(Z_1^T B Z_2)\big\}} \leq O(1)\frac{\mathrm{Tr}(\Sigma^4)}{\mathrm{Tr}^2(\Sigma^2)} + O(1)\frac{\mathrm{Tr}^4(\Sigma)}{\mathrm{Tr}^2(\Sigma^2)}\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big) = o(1).$$

This proves (8). □

Lemmas A.3-A.5 below are useful for proving the results under local alternatives.

Lemma A.3 *For any p-dimensional vectors X and μ,*
(1) $\Big\|\frac{X-\mu}{\|X-\mu\|} - \frac{X}{\|X\|}\Big\| \leq 2\frac{\|\mu\|}{\|X\|}$.
(2) $\Big\|\frac{X-\mu}{\|X-\mu\|} - \frac{X}{\|X\|} - \frac{1}{\|X\|}\big(I_p - \frac{XX^T}{\|X\|^2}\big)\mu\Big\| \leq c_2\frac{\|\mu\|^{1+\delta}}{\|X\|^{1+\delta}}$, *for all $0 < \delta < 1$, where c_2 is a constant that does not depend on X or μ.*

Proof. See Lemma 6.2 of Oja (2010).

Lemma A.4 *Let B be the matrix defined in Lemma 2.1. Assume condition (C3) holds, then $\lambda_{\max}(B) \leq \frac{2\lambda_{\max}(\Sigma)}{Tr(\Sigma)}(1+o(1))$.*

Proof. Let A_1 be the event defined as in the proof of Lemma 2.1. and α be an arbitrary unit length p-dimensional vector. Similarly as in the proof for Lemma 2.1,

$$\begin{aligned}
\lambda_{\max}(B) &= \max_{\|\alpha\|=1} E\Big(\frac{\alpha^T\Gamma U_1 U_1^T \Gamma^T \alpha}{U_1^T\Gamma^T\Gamma U_1}\Big) \\
&\leq \max_{\|\alpha\|=1} E\Big(\frac{\alpha^T\Gamma U_1 U_1^T \Gamma^T \alpha}{U_1^T\Gamma^T\Gamma U_1} I(A_1)\Big) + P(A_1^c) \\
&\leq \frac{\max_{\|\alpha\|=1}\Big\{\alpha^T\Gamma E(U_1U_1^T)\Gamma^T\alpha\Big\}}{\frac{1}{2p}\mathrm{Tr}(\Omega)} + 2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big)
\end{aligned}$$

Then it follows by condition (C3) that

$$\lambda_{\max}(B) \leq \frac{2\lambda_{\max}(\Sigma)}{\mathrm{Tr}(\Sigma)} + 2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big) = \frac{2\lambda_{\max}(\Sigma)}{\mathrm{Tr}(\Sigma)}(1+o(1)),$$

Lemma A.5 *Let A be the matrix defined in Theorem 2.3 and $D = E\Big\{\frac{1}{\|\epsilon_1\|^2}\Big(I_p -$*

$\frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\Big\}$, *then* $\lambda_{\max}(A) \le E(\|\epsilon_1\|^{-1})$ *and* $\lambda_{\max}(D) \le E(\|\epsilon_1\|^{-2})$. *Furthermore, if conditions (C3) and (C4) hold, then* $\lambda_{\min}(A) \ge \frac{1}{\sqrt{3}}E(\|\epsilon_1\|^{-1})(1-o(1))$.

Proof. We first note that the eigenvalues of the matrix $I_p - \frac{\epsilon\epsilon^T}{\|\epsilon\|^2}$ are all between zero and one. This leads to the upper bounds of $\lambda_{\max}(A)$ and $\lambda_{\max}(D)$. Let A_1 be the event defined as in the proof Lemma 2.1, then similarly as before,

$$
\begin{aligned}
\lambda_{\min}(A) &= \min_{\|\alpha\|=1} E\Big\{\frac{1}{\|\epsilon_1\|}\alpha^T\Big(I_p - \frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\alpha\Big\} \\
&\ge \min_{\|\alpha\|=1} E\Big\{\frac{1}{\|\epsilon_1\|}\alpha^T\Big(I_p - \frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\alpha I(A_1)\Big\} \\
&\ge \sqrt{\frac{2p}{3\mathrm{Tr}(\Omega)}}E(|R_1|^{-1}) \min_{\|\alpha\|=1} E\Big\{\alpha^T\Big(I_p - \frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\alpha I(A_1)\Big\} \\
&\ge \sqrt{\frac{2p}{3\mathrm{Tr}(\Omega)}}E(|R_1|^{-1})\Big[\min_{\|\alpha\|=1}\Big\{\alpha^T E\Big(I_p - \frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\alpha\Big\} - P(A_1^c)\Big] \\
&= \sqrt{\frac{2p}{3\mathrm{Tr}(\Omega)}}E(|R_1|^{-1})\Big[\lambda_{\min}(I_p - B) - P(A_1^c)\Big] \\
&= \sqrt{\frac{2p}{3\mathrm{Tr}(\Omega)}}E(|R_1|^{-1})\Big[1 - \lambda_{\max}(B) - 2c_1\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big)\Big] \\
&\ge \sqrt{\frac{2p}{3\mathrm{Tr}(\Omega)}}E(|R_1|^{-1})(1-o(1)), \qquad \text{(A.3)}
\end{aligned}
$$

where the second inequality follows because $\frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}$ and A_1 only depend on U_1, which is independent of R_1; the last inequality applies Lemma A.4, conditions (C3) and (C4).

Recall that $\|\epsilon_1\| = |R_1||U_1^T\Gamma^T\Gamma U_1|^{1/2}$. We note that

$$
\begin{aligned}
E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}\big] &\le E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}I(A_1)\big] + E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}I(A_1^c)\big] \\
&\le \sqrt{\frac{2p}{\mathrm{Tr}(\Omega)}} + 2c_1\lambda_{\min}^{-1}(\Omega)\exp\Big(-\frac{\mathrm{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\Big) \\
&\le \sqrt{\frac{2p}{\mathrm{Tr}(\Omega)}} + E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}\big]o(1),
\end{aligned}
$$

where the last inequality follows by applying condition (C3) and observing $E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}\big] \geq \lambda_{\max}^{-1}(\Omega)$. This implies

$$\sqrt{\frac{2p}{\mathrm{Tr}(\Omega)}} \geq E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}\big](1-o(1)). \tag{A.4}$$

Combining (A.3) and (A.4), we have

$$\lambda_{\min}(A) \geq \frac{1}{\sqrt{3}}E\big[|U_1^T\Gamma^T\Gamma U_1|^{-1}\big]E(|R_1|^{-1})(1-o(1)) = \frac{1}{\sqrt{3}}E(\|\epsilon\|^{-1})(1-o(1)).$$

Appendix 2: Proof of main theorems

In the sequel, we use c or C to denote generic positive constants, which may vary from line to line.

Proof of Theorem 2.2. Let $S_n^2 = Var(T_n) = \frac{n(n-1)}{2}\mathrm{Tr}(B^2) = \frac{n(n-1)}{2}E\big\{(Z_1^TZ_2)^2\big\}$. To apply the martingale central limit theorem (Hall and Heyde, 1980), it is sufficient to check two conditions:

$$\begin{aligned} S_n^{-4}\sum_{i=2}^n E(Y_i^4) &\rightarrow 0 \quad \text{as } n,p\rightarrow\infty, && \text{(A.5)}\\ S_n^{-2}V_n^2 &\rightarrow 1 \quad \text{in probability as } n,p\rightarrow\infty && \text{(A.6)} \end{aligned}$$

where $V_n^2 = \sum_{i=2}^n E(Y_i^2|Z_1,\ldots,Z_{i-1})$ and $Y_i = \sum_{j=1}^{i-1} Z_i^TZ_j$. To check (A.5), note that under H_0,

$$\begin{aligned} E(Y_i^4) &= E\Big\{\big(\sum_{j=1}^{i-1} Z_i^TZ_j\big)^4\Big\} = \sum_{j=1}^{i-1}E\big\{(Z_i^TZ_j)^4\big\} + 3\sum_{\substack{1\leq j,k\leq i-1\\ j\neq k}} E\Big\{(Z_i^TZ_j)^2(Z_i^TZ_k)^2\Big\}\\ &= (i-1)E\big\{(Z_2^TZ_2)^4\big\} + 3(i-1)(i-2)E\Big\{(Z_1^TZ_2)^2(Z_1^TZ_3)^2\Big\}. \end{aligned}$$

Hence, $\sum_{i=1}^n E(Y_i^4) \leq c\Big[n^2E\big\{(Z_1^TZ_2)^4\big\}+n^3E\big\{(Z_1^TZ_2)^2(Z_1^TZ_3)^2\big\}\Big] \leq cn^3E\big\{(Z_1^TZ_2)^4\big\}$ by Hölder's inequality. By Lemma 2.1, we have $E\big\{(Z_1^TZ_2)^4\big\} = o\Big(nE^2\big\{(Z_1^TZ_2)^2\big\}\Big)$.

Therefore, (A.5) holds.

To prove (A.6), it is sufficient to verify that $\frac{E(V_n^2-S_n^2)^2}{S_n^4} \to 0$ as $n, p \to \infty$. We write $V_n^2 = \sum_{i=2}^n V_{ni}$, where $V_{ni} = E(Y_i^2|Z_1, \ldots, Z_{i-1})$. We have

$$\begin{aligned} V_{ni} &= \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} E(Z_i^T Z_j Z_i^T Z_k|Z_1,\ldots,Z_{i-1}) = \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \mathrm{Tr}(Z_j Z_k^T B) = \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} Z_j^T B Z_k \\ &= 2\sum_{1\le j<k\le i-1} Z_j^T B Z_k + \sum_{j=1}^{i-1} Z_j^T B Z_j. \end{aligned}$$

If $j_1 \le k_1$ and $j_2 \le k_2$, then

$$\begin{aligned} & E\big(Z_{j_1}^T B Z_{k_1} Z_{j_2}^T B Z_{k_2}\big) \\ = \; & E\big((Z_1^T B Z_1)^2\big) I\{j_1 = k_1 = j_2 = k_2\} + E^2\big(Z_1^T B Z_1\big) I\{j_1 = k_1 \ne j_2 = k_2\} \\ & + E\big((Z_1^T B Z_2)^2\big) I\{j_1 = j_2, k_1 = k_2, j_1 < k_1\}. \end{aligned}$$

Therefore, for $i_1 \le i_2$,

$$\begin{aligned} & E\big(V_{ni_1} V_{ni_2}\big) \\ = \; & 4\sum_{1\le j<k\le i_1-1} E\big\{(Z_1^T B Z_2)^2\big\} + \sum_{j=1}^{i_1-1}\sum_{k=1}^{i_2-1} E^2\big(Z_1^T B Z_1\big) + \sum_{j=1}^{i_1-1}\big[E\big\{(Z_1^T B Z_1)^2\big\} - E^2(Z_1^T B Z_1)\big] \\ = \; & 2(i_1-1)(i_1-2)E\big\{(Z_1^T B Z_2)^2\big\} + (i_1-1)(i_2-1)E^2\big(Z_1^T B Z_1\big) + (i_1-1)Var(Z_1^T B Z_1). \end{aligned}$$

Consequently,

$$\begin{aligned} E(V_n^4) &= E\big\{\big(\sum_{i=2}^n V_{ni}\big)^2\big\} = 2\sum_{2\le i<j\le n} E(V_{ni}V_{nj}) + \sum_{i=2}^n E(V_{ni}^2) \\ &= 2\sum_{i=2}^n (i-1)(i-2)(2n-2i+1)E\big\{(Z_1^T B Z_2)^2\big\} + \sum_{i=2}^n (i-1)(2n-2i+1)Var(Z_1^T B Z_1) \\ &\quad + \Big\{n(n-1)E\big(Z_1^T B Z_1\big)/2\Big\}^2. \end{aligned}$$

Note that $E\big(Z_1^T BZ_1\big) = \mathrm{Tr}(B^2)$ and $S_n^2 = \frac{n(n-1)}{2}\mathrm{Tr}(B^2)$. Hence,

$$\begin{aligned}
& E(V_n^2 - S_n^2)^2 = E(V_n^4) - S_n^4 \\
= \;& 2\sum_{i=2}^{n}(i-1)(i-2)(2n-2i+1)E\big\{(Z_1^T BZ_2)^2\big\} + \sum_{i=2}^{n}(i-1)(2n-2i+1)Var(Z_1^T BZ_1) \\
\leq \;& c\Big[n^4 E\big\{(Z_1^T BZ_2)^2\big\} + n^3 E\big\{(Z_1^T BZ_1)^2\big\}\Big].
\end{aligned}$$

Hence, a sufficient condition for $S_n^{-4}E(V_n^2 - S_n^2)^2 \to 0$ is

$$\frac{n^4 E\big\{(Z_1^T BZ_2)^2\big\} + n^3 E\big\{(Z_1^T BZ_1)^2\big\}}{n^4 E^2\big\{(Z_1^T BZ_1)^2\big\}} \to 0.$$

This condition holds by Lemma 2.1. This finishes the proof of Theorem 2.2. $\square$

Derivation of (10). Let $\overline{Z}^* = (n-2)^{-1}\sum_{m=1}^{n} Z_m$ and

$$\begin{aligned}
A_1 &= \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\sum_{1\leq j\neq k\leq n}(Z_j - \overline{Z}^*)Z_j^T(Z_k - \overline{Z}^*)Z_k^T\Big\} \\
A_2 &= \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\sum_{1\leq j\neq k\leq n}(Z_j - \overline{Z}^*)Z_j^T(\overline{Z}^* - \overline{Z}_{(j,k)})Z_k^T\Big\}, \\
A_3 &= \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\sum_{1\leq j\neq k\leq n}(\overline{Z}^* - \overline{Z}_{(j,k)})Z_j^T(\overline{Z}^* - \overline{Z}_{(j,k)})Z_k^T\Big\}
\end{aligned}$$

Following (9), we have

$$\begin{aligned}
& \widehat{\mathrm{Tr}(B^2)} \\
= \;& \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\sum_{1\leq j\neq k\leq n}\big[(Z_j - \overline{Z}^*) + (\overline{Z}^* - \overline{Z}_{(j,k)})\big]Z_j^T\big[(Z_k - \overline{Z}^*)) + (\overline{Z}^* - \overline{Z}_{(j,k)})\big]Z_k^T\Big\} \\
= \;& A_1 + 2A_2 + A_3.
\end{aligned}$$

We next simplify A_1.

$$\begin{aligned} A_1 &= \{n(n-1)\}^{-1}\mathrm{Tr}\Big\{\big(\sum_{j=1}^n (Z_j-\overline{Z}^*)Z_j^T\big)^2\Big\} - \{n(n-1)\}^{-1}\sum_{j=1}^n \big[Z_j^T(Z_j-\overline{Z}^*)\big]^2 \\ &\quad -\{n(n-1)\}^{-1}\sum_{j=1}^n \big(1-2Z_j^T\overline{Z}^* + \overline{Z}^{*T}Z_jZ_j^T\overline{Z}^*\big) \\ &= \{n(n-1)\}^{-1}\mathrm{Tr}\{(\sum_{j=1}^n Z_jZ_j^T-(n-2)\overline{Z}^*\overline{Z}^{*T})^2\} \\ &\quad -\{n(n-1)\}^{-1}\{n-2(n-2)\left\|\overline{Z}^*\right\|^2+\overline{Z}^{*T}(\sum_{j=1}^n Z_jZ_j^T)\overline{Z}^*\}. \end{aligned}$$

Similarly, it follows that

$$\begin{aligned} A_2 &= -2\{(n-1)(n-2)\}^{-1}+\{n(n-1)(n-2)\}^{-1}\mathrm{Tr}\{(\sum_{j=1}^n Z_jZ_j^T)^2\} \\ &\quad -\{n(n-1)\}^{-1}\overline{Z}^{*T}(\sum_{j=1}^n Z_jZ_j^T)\overline{Z}^* \\ A_3 &= (n-4)\{(n-1)(n-2)^2\}^{-1}+2\{n(n-1)\}^{-1}||\overline{Z}^*||^2 \\ &\quad +\{n(n-1)(n-2)^2\}^{-1}\mathrm{Tr}\{(\sum_{j=1}^n Z_jZ_j^T)^2\}. \end{aligned}$$

Putting the above together, we get (10). □

Proof of Theorem 2.3. Under the local alternatives,

$$\begin{aligned} T_n &= \sum_{i=1}^n\sum_{\substack{j=1\\j<i}}^n \Big\{\frac{\epsilon_i}{\|\epsilon_i\|}+\Big(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}-\frac{\epsilon_i}{\|\epsilon_i\|}\Big)\Big\}^T\Big\{\frac{\epsilon_j}{\|\epsilon_j\|}+\Big(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}-\frac{\epsilon_j}{\|\epsilon_j\|}\Big)\Big\} \\ &= T_{n1}+T_{n2}+T_{n3}, \end{aligned}$$

where $T_{n1}=\sum_{i=1}^n\sum_{\substack{j=1\\j<i}}^n \frac{\epsilon_i^T\epsilon_j}{\|\epsilon_i\|\|\epsilon_j\|}$, $T_{n2}=\sum_{i=1}^n\sum_{\substack{j=1\\j\neq i}}^n \left(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}-\frac{\epsilon_i}{\|\epsilon_i\|}\right)^T\frac{\epsilon_j}{\|\epsilon_j\|}$, and $T_{n3}=\sum_{i=1}^n\sum_{\substack{j=1\\j<i}}^n \left(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}-\frac{\epsilon_i}{\|\epsilon_i\|}\right)^T\left(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}-\frac{\epsilon_j}{\|\epsilon_j\|}\right)$. By Theorem 2.2, $T_{n1}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)}\to N(0,1)$.

To analyze T_{n2}, we write $T_{n2} = T_{n21} + T_{n22}$, where $T_{n21} = \sum_{i<j} \left(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|} - \frac{\epsilon_i}{\|\epsilon_i\|} \right)^T \frac{\epsilon_j}{\|\epsilon_j\|}$ and $T_{n22} = \sum_{j<i} \left(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|} - \frac{\epsilon_i}{\|\epsilon_i\|} \right)^T \frac{\epsilon_j}{\|\epsilon_j\|}$. Note that $E(T_{n21}) = 0$, and

$$\begin{aligned}
E(T_{n21}^2) &= \sum_{i_1<j_1}\sum_{i_2<j_2} E\Big\{\Big(\frac{\epsilon_{i_1}+\mu}{\|\epsilon_{i_1}+\mu\|} - \frac{\epsilon_{i_1}}{\|\epsilon_{i_1}\|}\Big)^T \frac{\epsilon_{j_1}}{\|\epsilon_{j_1}\|}\frac{\epsilon_{j_2}^T}{\|\epsilon_{j_2}\|}\Big(\frac{\epsilon_{i_2}+\mu}{\|\epsilon_{i_2}+\mu\|} - \frac{\epsilon_{i_2}}{\|\epsilon_{i_2}\|}\Big)\Big\} \\
&= \sum_{i_1<j}\sum_{i_2<j} E\Big\{\Big(\frac{\epsilon_{i_1}+\mu}{\|\epsilon_{i_1}+\mu\|} - \frac{\epsilon_{i_1}}{\|\epsilon_{i_1}\|}\Big)^T B\Big(\frac{\epsilon_{i_2}+\mu}{\|\epsilon_{i_2}+\mu\|} - \frac{\epsilon_{i_2}}{\|\epsilon_{i_2}\|}\Big)\Big\} \\
&\le \lambda_{\max}(B)\sum_{i_1<j}\sum_{i_2<j} E\Big\{\frac{4\,\|\mu\|^2}{\|\epsilon_{i_1}\|\,\|\epsilon_{i_2}\|}\Big\} \\
&\le 2\frac{\lambda_{\max}(\Sigma)}{\mathrm{Tr}(\Sigma)}(1+o(1))\,\|\mu\|^2\Big\{\sum_{i<j} E(\|\epsilon_i\|^{-2}) + \sum_{i_1<j}\sum_{i_2<j} E^2(\|\epsilon_{i_1}\|^{-1})\Big\} \\
&\le O(n^3\,\|\mu\|^2)\frac{\lambda_{\max}(\Sigma)}{\mathrm{Tr}(\Sigma)} E(\|\epsilon\|^{-2}),
\end{aligned}$$

where the first inequality uses Lemma A.3, and the second inequality uses Lemma A.4. In the derivation in Lemma 2.1, we derived that $\mathrm{Tr}(B^2) \ge \frac{4}{9}\frac{\mathrm{Tr}(\Sigma^2)}{\mathrm{Tr}^2(\Sigma)}(1-o(1))$. Hence, it follows by condition (C5) that

$$\frac{E(T_{n21}^2)}{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)} \le O(n\,\|\mu\|^2)\frac{\lambda_{\max}(\Sigma)\mathrm{Tr}(\Sigma)}{\mathrm{Tr}^2(\Sigma)} E(\|\epsilon\|^{-2}) = o(1).$$

This implies $T_{n21}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)} = o_p(1)$. Similarly, $T_{n22}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)} = o_p(1)$.

Finally, we analyze T_{n3}. Denote

$$\begin{aligned}
T_{n31} &= \frac{n(n-1)}{2} E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)^T E\Big(\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}\Big), \\
T_{n32} &= \sum_{j\ne i} \mathrm{E}\Big(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}\Big)^T\Big\{\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|} - \frac{\epsilon_j}{\|\epsilon_j\|} - E\Big(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}\Big)\Big\}, \\
T_{n33} &= \sum_{j<i}\Big\{\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|} - \frac{\epsilon_i}{\|\epsilon_i\|} - E\Big(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}\Big)\Big\}^T\Big\{\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|} - \frac{\epsilon_j}{\|\epsilon_j\|} - E\Big(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}\Big)\Big\}.
\end{aligned}$$

Then it follows that

$$
\begin{aligned}
T_{n3} &= \sum_{i=1}^{n}\sum_{\substack{j=1\\ j<i}}^{n}\Big[E\Big(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}\Big)+\Big\{\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}-\frac{\epsilon_i}{\|\epsilon_i\|}-E\Big(\frac{\epsilon_i+\mu}{\|\epsilon_i+\mu\|}\Big)\Big\}\Big]^T \\
&\quad\times\Big[E\Big(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}\Big)+\Big\{\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}-\frac{\epsilon_j}{\|\epsilon_j\|}-E\Big(\frac{\epsilon_j+\mu}{\|\epsilon_j+\mu\|}\Big)\Big\}\Big] \\
&= T_{n31}+T_{n32}+T_{n33}.
\end{aligned}
$$

To analyze T_{n31}, by Lemma A.3 (2), we can write $E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)=-A\mu+E(Q_1)$, where the remainder term Q_1 satisfies $E(\|Q_1\|^2)\le c_3\,\|\mu\|^{2+2\delta}\,E\big(\,\|\epsilon_1\|^{-2-2\delta}\big)$ for all $0<\delta<1$, where c_3 is a constant that does not depend on ϵ_1 or μ. Hence,

$$
E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)^T E\Big(\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}\Big)=\mu^TA^2\mu-\mu^TAE(Q_1)-\mu^TAE(Q_2)+E(Q_1)^TE(Q_2).
$$

Note that the last three terms on the right-hand side of the above expression are bounded by

$$
\begin{aligned}
&2c_3^{1/2}\,\|\mu\|^{1+\delta}\,\big\|\mu^TA\big\|\cdot E^{1/2}\Big(\,\|\epsilon_i\|^{-2-2\delta}\Big)+c_3\,\|\mu\|^{2+2\delta}\,E\Big(\,\|\epsilon_i\|^{-2-2\delta}\Big) \\
\le\;& c\,\|\mu\|^2\,o\big(E^2(\|\epsilon_i\|^{-1})\big)=o(\mu^TA^2\mu),
\end{aligned}
$$

where the inequalities use Lemma A.5 and condition (C6). Therefore, $T_{n31}=\frac{n(n-1)}{2}\mu^TA^2\mu(1+o(1))$.

To evaluate T_{n32}, we observe that $E(T_{n32})=0$ and that

$$
\begin{aligned}
E(T_{n32}^2) &= O(n^3)E\Big(\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}\Big)^TE\Big\{\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big) \\
&\quad\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big)^T\Big\}E\Big(\frac{\epsilon_3+\mu}{\|\epsilon_3+\mu\|}\Big).
\end{aligned}
$$

Note that by Lemma A.3 (2),

$$\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)=-\Big\{\frac{1}{\|\epsilon_1\|}\Big(I_p-\frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)-A\Big\}\mu+(Q_1-E(Q_1)).$$

Applying the above decomposition, we obtain

$$\begin{aligned}
&\lambda_{\max}\Big[E\Big\{\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big)\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big)^T\Big\}\Big]\\
\leq\;& 2\|\mu\|^2\lambda_{\max}(D)+C\|\mu\|^{2+2\delta}E\Big(\|\epsilon_1\|^{-2-2\delta}\Big),
\end{aligned}$$

where $D=E\Big\{\frac{1}{\|\epsilon_1\|^2}\Big(I_p-\frac{\epsilon_1\epsilon_1^T}{\|\epsilon_1\|^2}\Big)\Big\}$. Therefore, by Lemma A.5, conditions (C5), (C6), and observing that $\mathrm{Tr}(B^2)\geq\frac{4\mathrm{Tr}(\Sigma^2)}{9\mathrm{Tr}^2(\Sigma)}(1-o(1))$, we have

$$\begin{aligned}
E(T_{n32}^2) &\leq O(n^3)\Big\{2\|\mu\|^2E\big(\|\epsilon_1\|^{-2}\big)+C\|\mu\|^{2+2\delta}E\big(\|\epsilon_i\|^{-2-2\delta}\big)\Big\}\Big\{\big\|\mu^TA\big\|^2+o(\mu^TA^2\mu)\Big\}\\
&\leq O(n^3)\Big\{2\|\mu\|^2E\big(\|\epsilon_1\|^{-2}\big)+C\|\mu\|^{2+2\delta}E\big(\|\epsilon_i\|^{-2-2\delta}\big)\Big\}\Big\{\|\mu\|^2E^2\big(\|\epsilon_1\|^{-1}\big)\Big\}\\
&= o(n^2\mathrm{Tr}(B^2)).
\end{aligned}$$

Therefore, $T_{n32}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)}=o_p(1)$.

To evaluate T_{n33}, we observe that $E(T_{n33})=0$ and that

$$\begin{aligned}
&E(T_{n33}^2)\\
=\;& n(n-1)E\Big[\Big\{\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big\}^T\Big\{\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}-\frac{\epsilon_2}{\|\epsilon_2\|}-E\Big(\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}\Big)\Big\}\\
&\Big\{\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}-\frac{\epsilon_2}{\|\epsilon_2\|}-E\Big(\frac{\epsilon_2+\mu}{\|\epsilon_2+\mu\|}\Big)\Big\}^T\Big\{\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big\}\Big]\\
\leq\;& n(n-1)\Big\{2\|\mu\|^2\lambda_{\max}(D)+C\|\mu\|^{2+2\delta}\times E\Big(\|\epsilon_1\|^{-2-2\delta}\Big)\Big\}\\
&\times E\Big[\Big\|\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}-\frac{\epsilon_1}{\|\epsilon_1\|}-E\Big(\frac{\epsilon_1+\mu}{\|\epsilon_1+\mu\|}\Big)\Big\|^2\Big]\\
\leq\;& O(n^2)\Big\{2\|\mu\|^2\lambda_{\max}(D)+C\|\mu\|^{2+2\delta}\times E\Big(\|\epsilon_1\|^{-2-2\delta}\Big)\Big\}^2=o(n^2\mathrm{Tr}(B^2))
\end{aligned}$$

by Lemma A.5, conditions (C5) and (C6). Therefore, $T_{n33}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)}=o_p(1)$.

Summarizing the above, $T_{n3}/\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)} = \frac{\frac{n(n-1)}{2}\mu^T A^2 \mu(1+o(1))}{\sqrt{\frac{n(n-1)}{2}\mathrm{Tr}(B^2)}} + o_p(1)$. This finishes the proof. $\square$